Hospital Peer Groups, Reliability, and Stabilization: Shrinking to the Right Mean

November 5, 2013

Presentation to the Federal Committee on Statistical Methodology

Washington, DC

Alex Bohl

MATHEMATICA
Policy Research

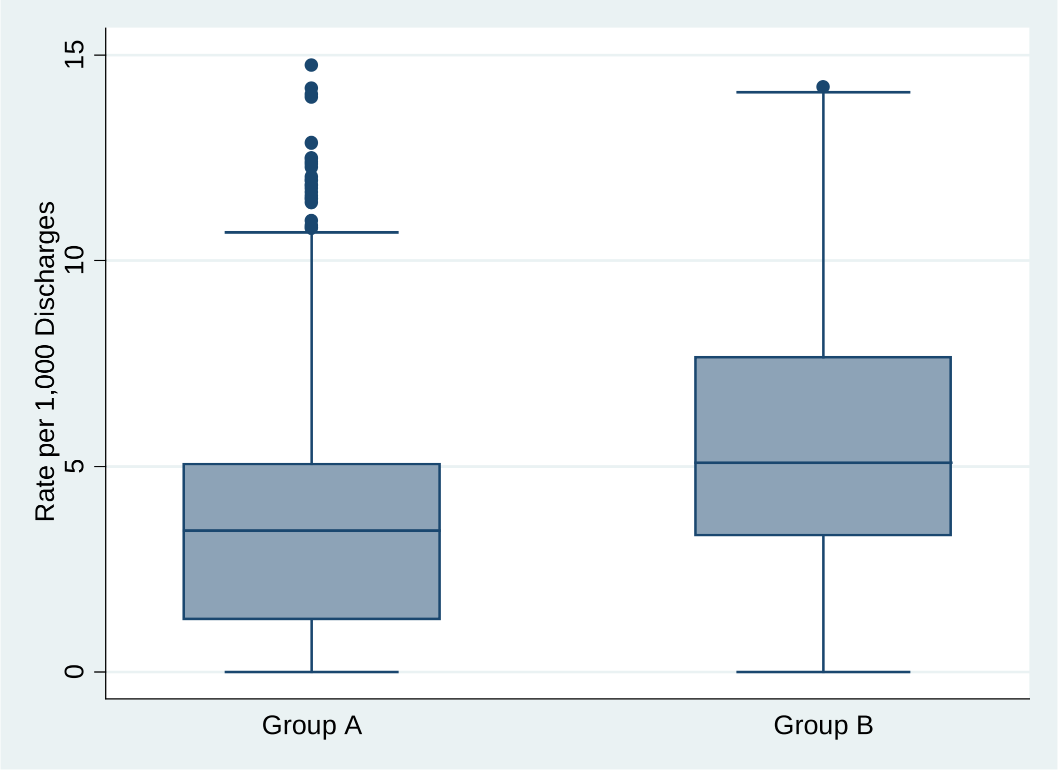
Agenda

- **Including Peer Groups in Hospital Comparisons**
  - **Rationale**
  - **Technical Approaches**
- **Empirical Example**
- **Challenges and Next Steps**

Stabilizing the Quality Indicators

- **Hospital Risk-Adjusted Rates (RARs) are often unstable**
  - **Small sample sizes**
  - **Rare events**

- **Smoothing stabilizes RARs by using information from the entire sample of hospitals**



What's the Correct Smoothing Target?

- Including hospital characteristics to create peer groups is controversial
  - Influences hospital ranking (Austin et al. 2004)
  - Changes the interpretation (Romano 2004)
- Volume is the most common characteristic considered
  - Strong empirical volume-outcome relationship for mortality (Silber et al. 2010)
- The ultimate choice of peer group
  - Needs conceptual and empirical backing
  - Depends on the outcome of interest
  - Should be precise

Technical Approaches to Peer Grouping

Hospital characteristics can enter risk- or reliability-adjustment models (or both)

- **Risk-Adjustment Model**
  - **Peer group fixed effects, and/or**
- **Reliability-Adjustment Model**
  - **One-part or unified: Smooth to peer group rates**
    - Peer group random effects
    - With or without risk adjustment for hospital-level factors
  - **Two-part shrinkage model: Standardize to peer group rate**
    - Estimate reliability as signal-to-noise ratio
    - Smooth to the peer group target

Illustrative Example

- **Aim**: Incorporate peer group targets into the AHRQ QI model
- **Peer grouping:** Teaching vs. Non-Teaching Affiliation
- **Measure:** PSI 12 (Postoperative Pulmonary Embolism or Deep Vein Thrombosis Rate)
- **Approach:**
  - Base case: Current QI methodology
  - Alternative: Two-part approach smoothing to teaching peer group target rates
- **Evaluation criteria:**
  - Change in reliability (signal variance/total variance)
  - Correlation of hospital ranking across approaches
  - Proportion of hospitals moving above/below national average

Methods

- **Calculate reliability weights and shrinkage targets for two scenarios:**
  - **Base Case (“Overall”)**
  - **Alternative (“Peer Group”)**
- **Reliability weights vary for each approach**
  - **Recalculate signal and noise**
- **Changes in smoothed rate estimates is therefore a function of**
  - **The new shrinkage target**
  - **The change in reliability weight**

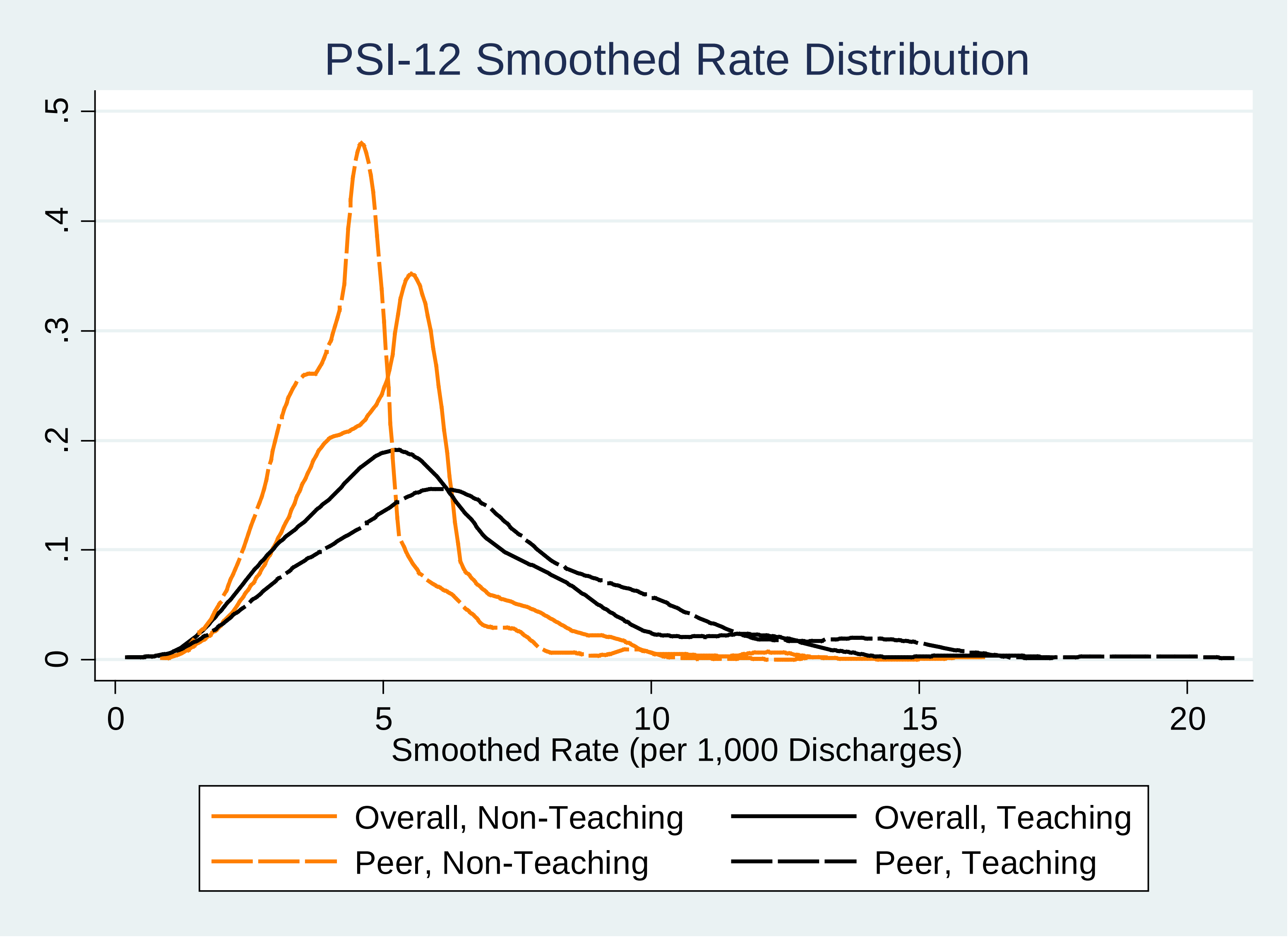
Descriptive Statistics: PSI 12 (DVT/PE)

	Overall	Non-Teaching	Teaching
Hospitals (n)	1,264	944	320
Denominator (mean)	4,605	3,056	9,177
Observed Rate	5.81	4.80	6.81
Expected Rate	5.81	5.52	6.11
Risk-Adjusted Rate	5.81	5.06	6.48

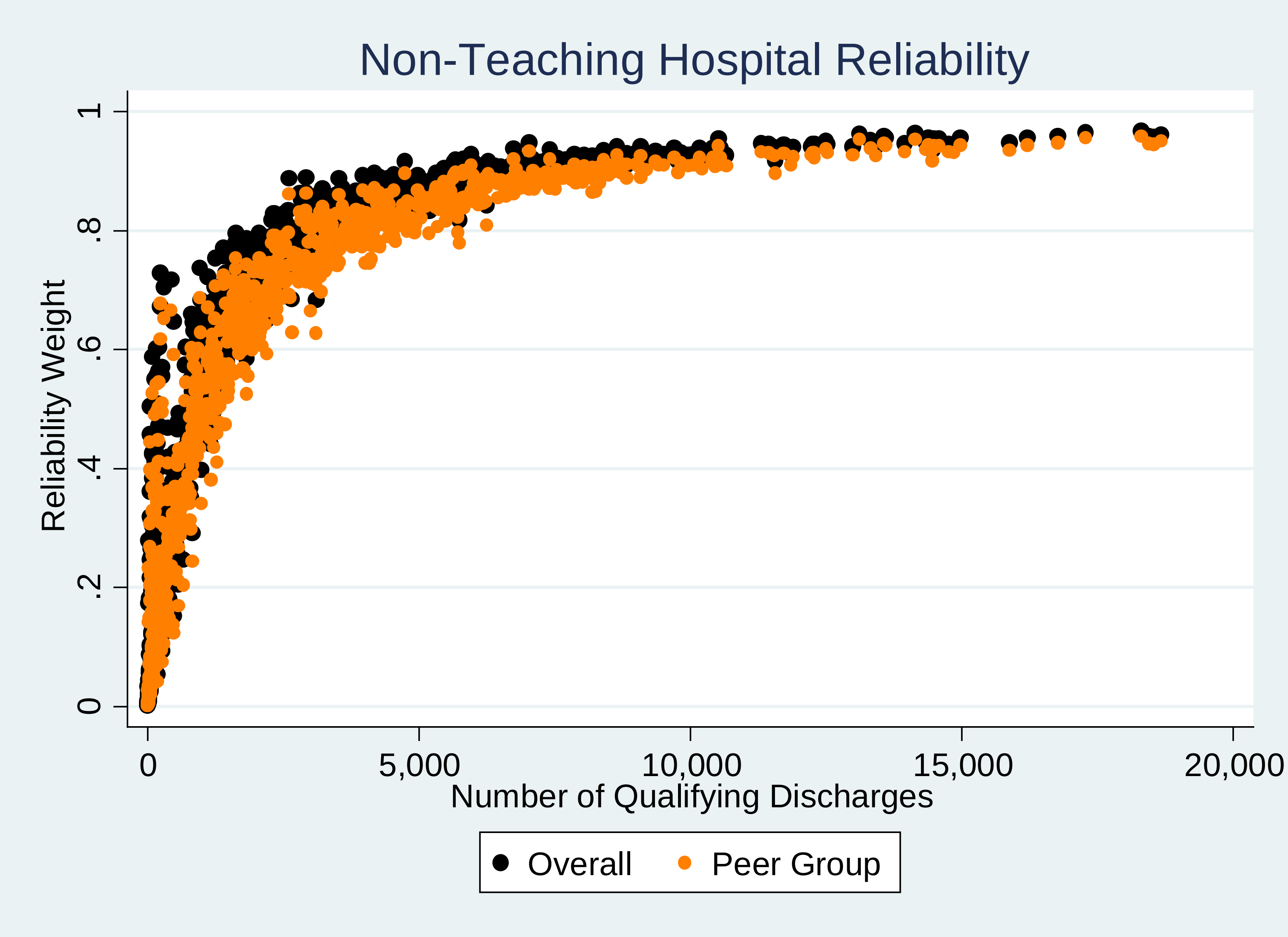
- **Rates have units per 1,000 discharges**
- **Random sample of hospitals from 12 states with Healthcare Cost and Utilization Project (HCUP) State Inpatient Databases (SIDs), 2009 and 2010***

* We would like to thank the HCUP Partners from the following states: AR, AZ, CA, FL, IA, KY, MA, MD, NE, NJ, NY, WA (http://www.hcup-us.ahrq.gov/partners.jsp)

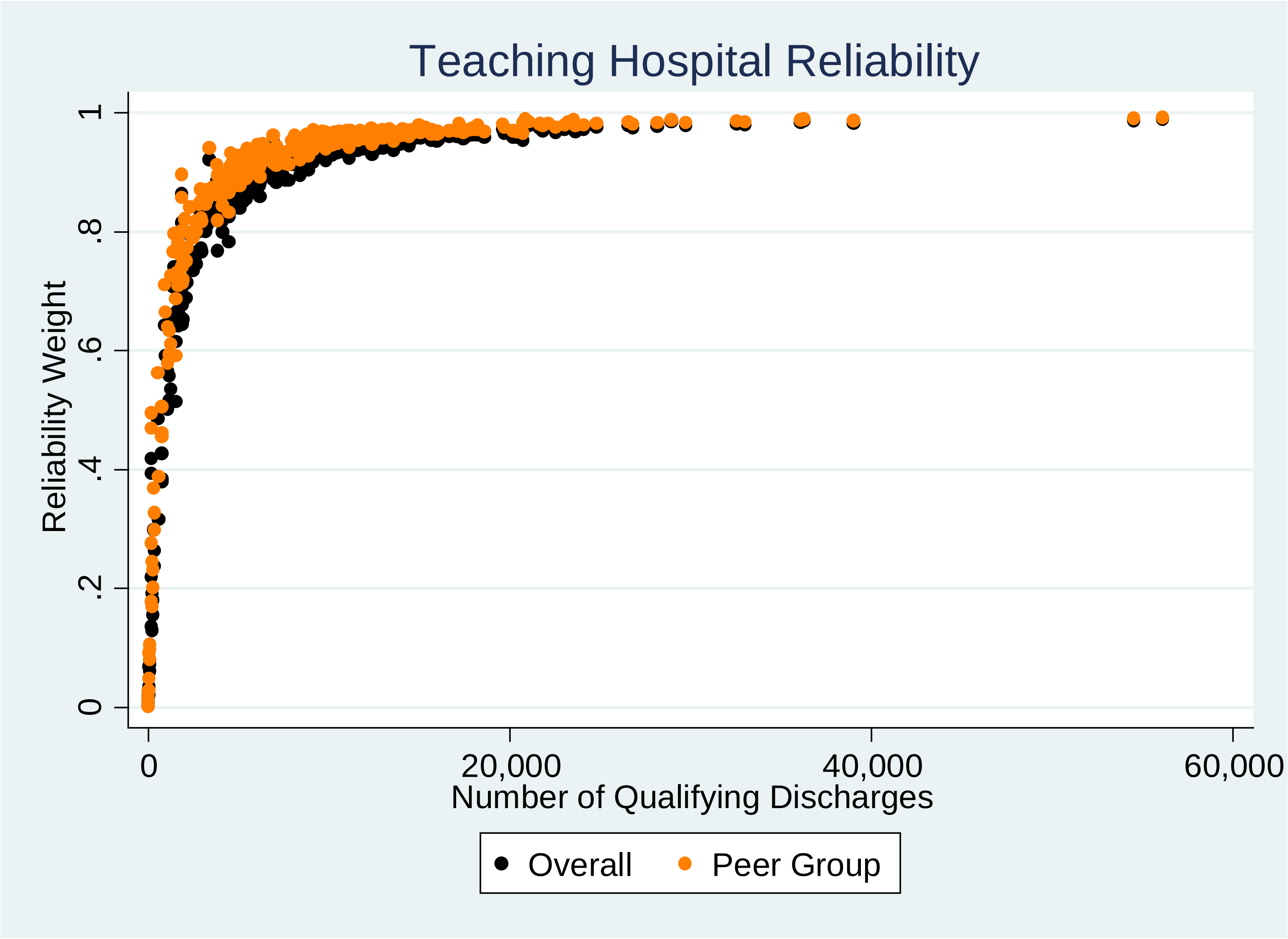
Smoothed Rate Distribution



Reliability Estimates – Non-Teaching Hospitals

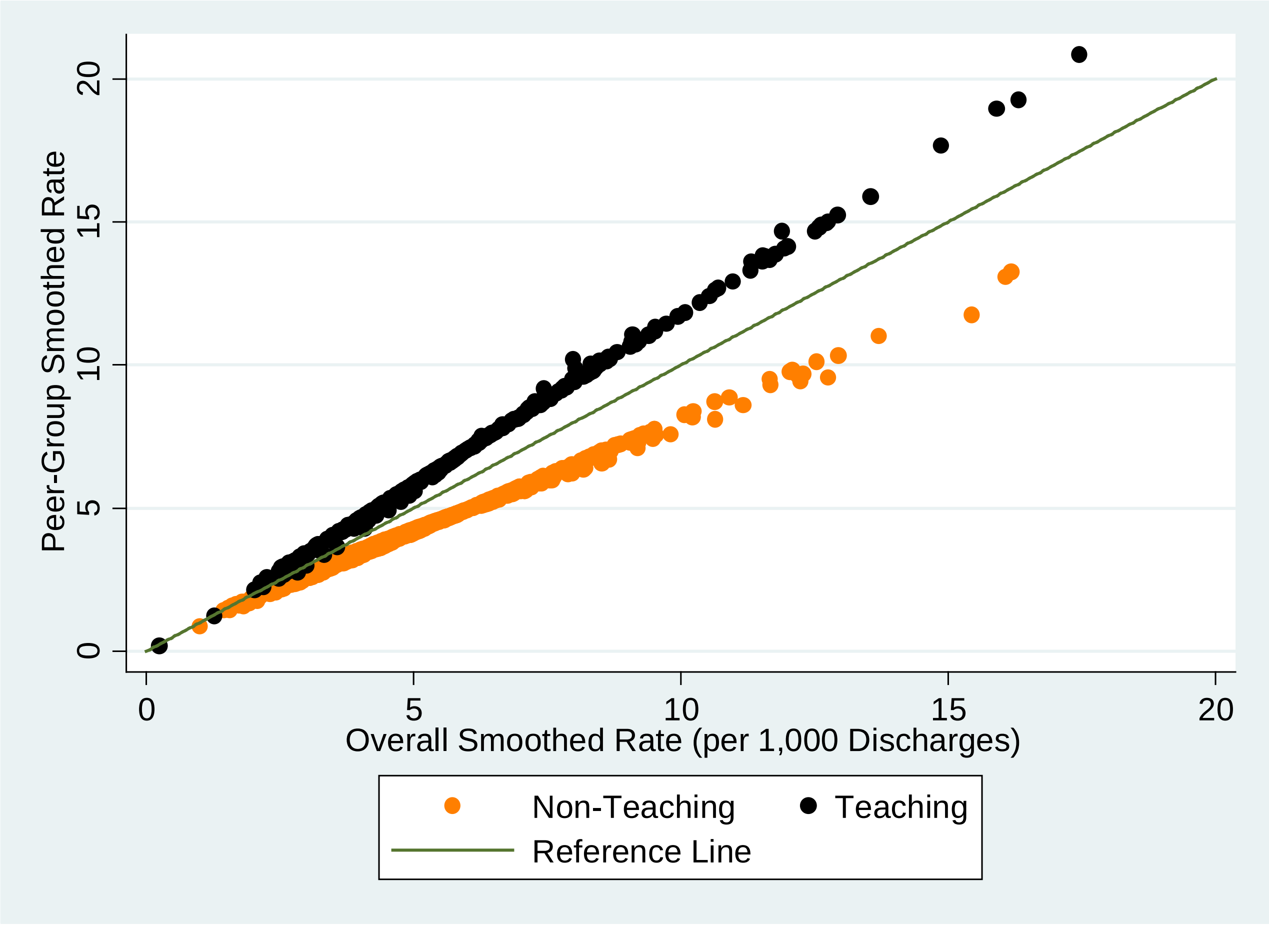


Reliability Estimates – Teaching Hospitals



MATHEMATICA Policy Research

Smoothed Rates



- **Teaching hospitals: 18% move above national average**
- **Non-teaching hospitals: 15% move below national average**

Summary

- **Using peer group targets changes ranking of smoothed rates**
  - **Teaching: 18% move above national average**
  - **Non-teaching: 15% move below national average**
  - **Rank sum correlation of 0.91**
- **Peer grouping changes the variability in PSI 12 distribution through reliability weights**
  - **Teaching: Increased variability**
  - **Non-teaching: Decreased variability**

Challenges and Limitations

Practical, Conceptual, and Technical Questions Remain

- What happens for hospitals on the boundary?
  - For example: volume, disproportionate share percentages, or nurse staffing ratios
- What about more precise subgroups?
  - Major versus minor teaching status
  - Subdividing non-teaching hospitals further
- What happens for small peer groups (e.g., two hospitals)?
- How do we handle hospitals missing peer group information?

Contact Information

- **Alex Bohl**
- **Mathematica Policy Research**
- **abohl@mathematica-mpr.com**
- **(617) 301-8996**

References

- **Austin et al. Impact of the choice of benchmark on the conclusions of hospital report cards. *American Heart Journal,* 148(6); 2004.**

- **Romano, P.S. Peer group benchmarks are not appropriate for health care quality report cards. *American Heart Journal,* 148(6); 2004.**

- **Silber et al. The Hospital Compare Mortality Model and the Volume-Outcome Relationship. *Health Services Research,* 45(5); 2010.**

Appendix: Estimating Noise

By the law of total variance:

$$\begin{aligned} Var(\epsilon_h) &= E\left\{Var\left(RAR_h - \theta_h|\theta_h\right)\right\} + Var\left\{E\left(RAR_h - \theta_h|\theta_h\right)\right\} \\ &= E\left\{Var\left(RAR_h|\theta_h\right)\right\} + \\ &\qquad E\left\{Var\left(\theta_h|\theta_h\right)\right\} + Var\left\{E\left(RAR_h - \theta_h|\theta_h\right)\right\} \end{aligned}$$

The last two terms drop out.

$$\begin{aligned} Var(\epsilon_h) &= E\left\{Var\left(RAR_h|\theta_h\right)\right\} \\ &= E\left\{Var\left(\overline{Y} \cdot \frac{O_h}{E_h}\right)\right\} \\ \hat{\sigma}_h^2 &= \left(\frac{\overline{Y}}{n_h \cdot E_h}\right)^2 \sum_{i \in A_h} \hat{Y}_i\left(1 - \hat{Y}_i\right) \end{aligned}$$

MATHEMATICA Policy Research

Appendix: Estimating Signal

Signal variance is the total variance $Var(RAR_h)$ minus the noise variance $Var(\epsilon_h)$. Note that:

$$E\left\{(RAR_h - \mu)^2 - \hat{\sigma}_h^2\right\} = Var\left(\theta_h\right)$$

Using this relation we have that:

$$\begin{aligned} Var(\theta_h) &= Var(RAR_h) - E(\hat{\sigma}_h^2) \\ \hat{\tau}^2 &= \frac{1}{H-1}\sum_h \left\{\left(RAR_h - \overline{RAR}\right)^2 - \hat{\sigma}_h^2\right\} \end{aligned}$$

MATHEMATICA Policy Research

Appendix: Estimating Reliability

We have assumed a simple linear regression which has a known solution found using the least-squares estimate or the maximum likelihood estimate: (MLE)

$$\theta_h - \mu \quad = \quad \lambda_h \cdot (RAR_h - \mu) + \omega_h$$

The MLE is given by:

$$\hat{\lambda}_h = \frac{Cov(\theta_h, RAR_h)}{Var(RAR_h)} = \frac{Var(\theta_h)}{Var(\theta_h) + Var(\epsilon_h)} = \frac{\tau^2}{\tau^2 + \sigma_h^2}$$

Use the relation $RAR_h = \theta_h + \epsilon_h$ to get the numerator result that $Cov(\theta_h, RAR_h) = Var(\theta_h)$.

Future Considerations

- **Multilevel random effects**
  - **Cross-classification of groups**
- **Incorporating peer groups (or different peer groups) into the risk-adjustment model**
- **Exploring the impact of historical priors, or priors defined outside the analytic population**
- **Application to patient safety indicators**
  - **Lower event rates**
  - **No consistent relationship with characteristics**

Conclusions

- **Whether to shrink to peer group means depends on**
  - **Empirical evidence**
  - **Conceptual background**
  - **Precise peer group classification**
  - **Desired interpretation or use**